DOCUMENT RESUME

ED 167 086 IR 006 761

AUTHOR Cound, William T.
TITLE To the Pacific: An Exploration of Computer-Based Reference Services.
PUB DATE 2 Dec 77
NOTE 47p.; Independent study, University of Denver Graduate School of Librarianship

EDRS PRICE MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS Cost Effectiveness; *Data Bases; Efficiency; Graphs; Relevance (Information Retrieval); *Search Strategies; Statistical Data
IDENTIFIERS *Aluminum Industry; *Computer Based Reference Services

ABSTRACT

In contrast to the usual approach in computer-based reference services, which is to go into a specific data base to retrieve citations to material on a specific, narrowly-defined topic, this report demonstrates how such services could be useful in a broad approach to a complex subject, using an investigation of trends in the world aluminum industry as a case in point. The user's request was broken down into three questions; eight of the 75 data bases available were finally used. A controlled, phased exploration of these data bases was conducted. At the conclusion of each phase, the recovered citations were cataloged, analyzed, and compared. Tentative conclusions were drawn, and the exploration redirected if indicated. The approach was successful in that it produced a wealth and variety of information, was cost-effective if one considers what a librarian would have to do to produce the same results using manual search methods, and seems cost-beneficial from the point of view of efficiency, precision, and effectiveness. Appendices present the search histories of the study by question and data base, the respective times online for the searches, a description of the data bases, descriptive statistics of the results, a cost-comparison of man and machine, tables of bauxite producers and aluminum companies, and a consideration of the precision obtained. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED167086

UNIVERSITY OF DENVER
GRADUATE SCHOOL OF LIBRARIANSHIP

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TO THE PACIFIC

AN EXPLORATION OF COMPUTER-BASED REFERENCE SERVICES

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William T. Cound

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

WILLIAM T. COUND
97-499.1 - Independent Study
DENVER, COLORADO
2 December, 1977

IR006761

TABLE OF CONTENTS

The report

About aluminum.....2

Distinction in the Approaches.....4

The Three Questions.....4

The Method.....5

Conclusions Reached.....7

Observations made and Lessons learned.....10

Annexes:

A: Search Histories

B: The Dialog and Orbit Data Bases Used - A viewpoint

C: Descriptive Statistics of the Results

D: An Example....of Titles Retrieved....

E: Bauxite Producers

F: Aluminum Companies

Computer-based reference services are a new and powerful tool to support research. They offer a means to cope with the rapidly expanding and evolving subjects and fast-moving events of concern to corporations, institutions, and individuals. They have promise of accomplishing the classic mission of the librarian effectively -- providing the patron with the information he wants at the minimum cost in time and with the greatest relevance.

Computer-based reference services are expensive, and their use is generally limited to the search for very specific answers because of this cost. But -- they have the potential of being cost-effective in gathering pertinent citations on broad and loosely defined subjects. It is these kinds of questions that so often require study but are avoided because of the tedious and time-consuming process of manual search and retrieval. Where time is money computer-based reference searches should be pursued. After all, these services are but the electronic expression of the numerous indexing and abstracting services with which we are familiar and which we use when the questions are pressing enough to make the effort. The answers are there; the question is how to find them.

This is a report of an experiment I conducted to find a way. I have a friend, who is a doctoral candidate. His dissertation is on trends in the world aluminum industry. During the winter quarter of 1977 I was taking a course in computer-based reference services. Enthusiastic about them I persuaded him to try them in his research. His disappointment with the result was a challenge to a fledgling librarian. I didn't know how to do it, but I was sure there was a way. After all, scholars have been searching the literature for generations, and in the past fifty years or so they have been materially aided by the growth of indexing and abstracting services.

Reviewing the search he had made and my own attempts I realized that my approach was too narrow and confined to the most obvious data base. I did not set out to prove anything or to exhaust the subject. Like Lewis and Clark's expedition to the Pacific I followed but a few trails and noted what I saw.

The purpose of this presentation is to report on the exploration of computer-based reference services to gather citations from a number of relevant data bases on a broad and complex subject.

By reporting the method used and the lessons learned I hope to encourage others to tackle other, broad, and complex subjects using these services and open a new and fruitful use of these services.

To give the exploration direction I used my friend's hypothesis:

> During the past decade there have been dramatic changes in the ownership patterns of world mineral productive capacity. This project will analyze these changes, summarize significant events, and suggest economic and institutional policies needed to reduce the risk of detrimental effects on the United States.
> Scope: worldwide sources of aluminum, alumina, and bauxite; ownership and control of mines, refining, and smelting facilities.[1]

---

1. S. B. Rogers, doctoral candidate in mining economics, Colorado School of Mines, Golden, Colorado.

Discussions with him before I began the exploration disclosed nongeological factors affecting supply and demand: potential cartelization of mineral producers; commodity agreements between producers and consumers; changes in the location and ownership of the total process stream; political disruption within countries possessing the mineral bauxite; U.S. stockpiling; and, significant breakthroughs in the technology of production of aluminum from bauxite and other minerals.

- About Aluminum -

Aluminum is the most plentiful metal in the earth's crust. It is not found, however, in its pure, metallic state. Aluminum is strong, light weight, and rust proof. It is easily worked, heat reflecting, and a good conductor of electricity. Aluminum is considered a strategic commodity, although its military importance has declined with the advances made in other light metals. Aluminum today is in a highly competitive situation with other materials, such as iron and steel for structures, copper for electrical uses, and plastics and paper for packaging and containers.[1]

Most aluminum produced is derived from the mineral bauxite. The bauxite is usually mined from open pits and beneficiated near the mine, that is, it is crushed and washed to remove silicon and dried to reduce weight and bulk for transportation. The mining and beneficiating of bauxite is a low cost operation. The capital investment is low and most of the equipment is movable so that it can be saved when the mining is discontinued for any reason.

Bauxite is refined into alumina (Al_2O_3) by the Bayer process, first developed in 1888. Refining reduces the bulk and weight significantly, on the order of four or six to one. Powdered bauxite is dissolved in a caustic soda (NaOH) under heat and pressure. The impurities settle out as "red mud," and the aluminum combines with other elements to form sodium aluminum oxide ($NaAlO_2$). This is precipitated in cooling towers as sodium hydroxide (NaOH) and aluminum trihydrate ($AlOH_3$). The latter is calcined in ovens, producing alumina. The capital investment in bauxite refining runs on the order of 150 dollars per ton of capacity per year.

The Hall process, developed in 1886, reduces alumina to aluminum by electrolysis in a cryolitic (Na_3AlF_6) bath, consuming carbon in the electrodes, to form metallic aluminum and carbon dioxide. Two pounds of alumina produce one pound of pig. The pig is subsequently refined into commercially pure aluminum or used with other metals to form a number of important alloys. Investment in this stage of production is high, six to seven times that of steel, typically 1300 dollars per ton of capacity per year for a 20,000 ton a year plant. Economy of scale reduces this significantly, to about 700 dollars a ton for a 200,000 ton a year plant. Electrical consumption is high, about ten kilowatt hours per pound of ingot. The carbon consumed in the anodes is another significant cost factor, about .625 pounds per pound of pig. The labor at this stage of production is highly skilled and expensive.

The technology for producing aluminum is well-known and highly developed. Production is an international affair: reduction of bauxite to alumina in the country of origin is common, although beneficiated bauxite is shipped to Gulf Coast states for processing from some of the Caribbean islands. For long

---

1. Britannica 3, Macropedia, Vol 1, p. 281.

hauls transportation costs are significantly reduced by reducing the bauxite to alumina in-country. The final stage of production is almost entirely confined to the advanced countries for a number of reasons. Cost of handling and shipping the finished aluminum or fabricated goods is high, so smelters are located close to the final markets. Cheap and abundant electrical power is necessary; skilled labor pools are needed; and the political and financial stability of advanced countries protect the enormous investments that go into the final stage of production.

The aluminum industry has long been favored by governmental policy: war-built plants, financial assistance, tax privileges, publicly financed power, tariffs, and other controls. The aluminum industry is vertically integrated. There are six major companies: ALCOA (U.S.), Alcan (Canada), Kaiser (U.S.), Reynolds (U.S.), Pechiney (France), and Alasuisse (Switzerland). Some other large metals companies are diversifying into aluminum. There are a number of smaller companies with a special share in the market. (Annex F.)

Bauxite (from Les Baux, France, where it was first found in large quantity) is the mineral of choice. Bauxite was probably formed by the leaching of aluminum compounds from igneous rock which probably accounts for its prevalence in tropical countries whose torrential rains made this possible. Bauxite is a mixture of several compounds, elements, and conglomerations (chiefly aluminum trihydrate, titanium, iron, and silica oxide). The amount of silica in the mineral is significant. Above 7% too much of the aluminum combines with it in the Bayer process and is lost or requires a second, costly recovery process.

The bauxite-producing countries have formed the International Bauxite Association to work towards more favorable prices and other conditions. They could, perhaps, dictate price and supply (like OPEC in oil) if they could only control the mineral. They could force the great producers to put more of the production process in-country, improving their economies and upgrading their work force. The producers resist this, not only because of significantly higher transportation costs but because of the risk to their heavy investment, and the cost of upgrading local labor to the skills required. The concern is real enough. In 1974 Jamaica took several steps giving substance to these fears: it increased royalties, added an export tax, and tied the price of bauxite to the price of primary aluminum in the country of production. It also maneuvered to acquire a controlling interest in the producers' local subsidiaries.

The situation is not like oil's, however. Bauxite is relatively plentiful to demand. Demand has softened considerably with the worldwide recession. Finished aluminum must compete with a number of other commodities. A few cents difference per pound is a real competitive edge. There are also a number of other minerals, in ready supply, that could supplant bauxite. On-going research and development in these minerals could bring costs in line with higher-priced bauxite. Control of bauxite by an international cartel or threat of governmental action might lead to location decisions by the major companies that would permanently destroy the bauxite market.[1] (Annex F, 1974 worldwide production.)

The other minerals of interest are: alunite (used for years by the Soviet Union that is willing to pay a price for independence from foreign suppliers);

---

1. Brubaker, Sterling. Trends in the World Aluminum Industry. Baltimore, Johns Hopkins Press, c1967. Particularly, Chapter 4 (basic technology), 5 (structure of the international industry), and 8 (technological prospects).

high alumina clays such as kaolin (found in the southeastern United States and already developed for the ceramics industry); anothosite; nepheline syenite; dawsonite; saprolite; aluminum phosphate; aluminous slate, and oil shale. Coal mine tailings and fly ash, two industrial waste products that are now a nuisance, and recycling aluminum waste are valuable sources whose costs are offset to some extent by pollution controls. The 1974 Jamaican actions increased the price sixfold to its customers. The formation of the International Bauxite Association has made producers uneasy. All these have made other sources besides bauxite attractive.[1]

- Distinctions in the Approaches -

The usual approach is to go into a specific data base to retrieve citations to material on a specific, narrowly defined topic, specifying as many characteristics of the topic as possible in the search strategy in order to exclude irrelevant material and obtain results with the minimum time on line.

The broad approach to a complex subject (used in this experiment) is to select every data base that might touch on some aspect of the subject, chosing only those characteristics that define the subject areas in order to gather in as much relevant material that will, in sum, define the many dimensions of the subject and adumbrate the directions in which it is evolving.

To illustrate: a company interested in whether it will be duplicating research to recover rare elements from the Bayer process would search World Aluminum Abstracts (WAA) with specific subject codes and terms related to the Bayer process. In this experiment my client is interested in everything that is going on in research that might affect aluminum production and costs. So, seven data bases are explored using the broadest terms associated with the concept of research and the mineral bauxite and the product alumina.

The distinctions made above are precise; in practice, of course, they would tend to blur and merge as the specific search was broadened.

- The Three Questions -

Discussions with my friend and reading several books he recommended made it apparent that I needed to develop some very precise questions to guide the search effectively. They had to be few in number to be workable; they had to be mutually exclusive; but, collectively, they had to cover all the hypothesis expressed or implied -- what I came to term as the "big question."

1. Where are the significant deposits of bauxite in the world today, and what, if any, changes have taken place in their distribution, the past ten years?

2. What, if any, experiments and research promise technology to produce alumimum from another mineral, other than bauxite, commercially?

3. What, if anything, is happening to the ownership or control of bauxite, alumina, or aluminum producing capacity?

It may look, at first, as if this is nothing but a group of specific searches, but these questions could not have been formulated without a very

---

1. United States. Senate. Permanent Subcommitee on Investigations of the Committee on Government Operations. "Materials shortages: aluminum." Committee Print. Washington, GPO, 1974.

specific knowledge of the data bases and the subject. Furthermore, they would not be coordinated -- lacking the integrating scheme to close over the gaps, identify redundancies, and direct the overall search properly.

The usual approach to computer-based reference services might arrive at the same results by fits and starts. It is questionable whether the researchers would ever broach the specific series of questions that would elicit the same, total response as this search directed over a broad front. In any event, they would not occur together, and the broad view and insights gained from conducting a multidirectional search in a short space of time would be lost.

- The Method -

A controlled, phased exploration of the pertinent data bases was conducted. At the conclusion of each phase the recovered citations were cataloged, analyzed, and compared. Tentative conclusions were drawn, and the exploration redirected if indicated. The method is given in some detail for those who might want to follow it in the exploration of some other broad, complex subject.

The first step was to research the aluminum industry to gain an understanding of its processes, organization, economics, and trends, and above all of its terminology. This research was useful in selecting data bases, formulating the search strategies, recognizing significant material while on line, and evaluating the "catch" after the search. (In effect I became my own client.) This research was broad but shallow but in far greater depth than the client interview which should proceed a specific search.

There were 45 data bases available at the computer access laboratory of the University of Denver at the time of this experiment. Nine were tentatively selected, but American Business Index/Inform was dropped after carefully reviewing its User Manual. As stated there, it is functionally oriented and is not product or company specific. Furthermore, it lacks a structured vocabulary and the chance of "snagging" something of value by an indirect approach was not promising. (A description of the eight chosen, drawn from the User Manuals and Data Base Summaries, is found at Annex B, p. B-2-1.)

Quick reference tables were then developed for use on-line. These lists of key words, category codes, and canned output commands helped refine the three questions further and to draft the search strategies. They were useful, on-line, particularly in creating the confidence needed to combat the pressure of the clock running off the "dollars."

Each data base was then related to one or more of the three questions. An initial search strategy was written for each data base, proceeding methodically from one of the questions to the next. These were not followed slavishly on-line, but it is easy to modify a written plan as circumstances on-line dictate. The attempt to improvise totally new strategies on-line was dismal and commend this practice for the confidence and directness written strategies provide.

The conduct of searches were guided by some predetermined rules which we[1] abandoned. They were too rigid. We followed where the responses led and cut off the search when it was unproductive. The terminal was connected to an auxiliary, computer-driven typewriter which made a complete record of each search. (The history of each search is found in Annex A.)

---

1. Professor James K. Foyle participated in the conduct of the searches. Although I am responsible for the study and its conclusions, only his guidance and encouragement made it possible.

After the on-line session each retrieval was transcribed on a color-coded, four-by-six inch card.

1858179 Revere Copper & Brass to sell Jamaican bauxite properties to government

(alumina + bauxite) * (ownership EC codes)

H F&S Index

[salmon-colored for Question 3]

The accession number (for identification) of the retrieval appears in the upper left-hand corner with the title to the right. An abbreviated statement of the logic used to obtain the retrieval appears in the center. The data base used in the search is shown in the lower-right. A pencilled evaluation of the retrieval (a hit) appears in the lower-left.

The actual transcription of the cards was useful; copying the information made me totally familiar with the recoveries. While such a 'tedious' task is unnecessary if the results need not be counted, sorted, categorized, and compared, simply handing the printout to the researcher will not do even when the approach is to support actual work, not an experiment. The retrievals need to be studied, evaluated, and analyzed by the librarian to guide follow-on searches.

Each retrieval was analyzed in terms of the question it addressed and the logic that recovered it. Retrievals were rated: hit, near miss, or clean miss. A hit was recorded if the retrieval pertained to the question addressed and responded to the logic. Near misses were counted when a retrieval missed the question addressed but provided an answer to another question or to the "big question." Like near misses in a fire fight, these retrievals contributed to the total effect -- hence their name. Some of the near misses could prove more useful to the client than many of the direct hits.

A question that can not be answered, due to the limits of the exploration, is whether near misses responding to the other two questions than the one addressed would have been retrieved if those questions had been broadened. All I can say is that there were no redundancies retrieved between two questions.

Was it or any other retrieval useful? There are really two judgments to be made about each retrieval. The first is the precision: a hit or a miss. The second is utility: does it help the client. The answers must be closely related if the search is to prove worthwhile. Still the two judgments are independent of one another. The first is a function of the question asked; the latter is a function of what the client already knows and what the retrieval contributes in the way of new knowledge or insights. My experiment confined itself to the first question. Only my friend's research, using the citations I supplied him, will answer the latter.

An arbitrary standard of precision (hits per retrieval) of 75% was established to investigate the less fruitful searches for reasons. (Annex B, p. B-3-1, contains a discussion.)

- Conclusions Reached -

The experiment led to a number of insights which I have labeled "conclusions." But, the experiment was only an exploration, and no conclusion can be hard and fast. (I can show you how I got to the Pacific and what I learned on the way.)

This disclaimer should be kept in mind by those who wish to use the experiment as a case study or to prove a point. Several sets were drastically reduced in size before they were printed, on-line, due to limitations of time and money. Some searches were explored well beyond the point necessary where -- in practice -- one would dump the set off-line for printing at far lower cost. In short, the method proved to be more cost-effective and efficient than the statistics would indicate. The search in SSIE could only be performed at a greater cost off-line. The only other way to obtain the information is to send a custom search request to the Smithsonian Science Information Exchange -- for a minimum fee of sixty dollars. The on-line search pulled the names and addresses of investigators for the client to correspond with at much less cost. The time on-line with Predicast's International Annual Time Series (Annex A, p. A-6-3) was spent in printing out statistics for 32 countries that could have been printed, off-line, at a fraction of the cost.

The most important conclusion is that this approach works. It is possible to use computer-based reference services for broad and complex subjects as well as narrowly defined topics. The method outlined is a start.

The practicality of the approach can be questioned considering the amount of time the librarian must spend on preliminary research and preparation. Various aspects of this point are discussed elsewhere below.

In any reasonable cost accounting the time of the researcher and the librarian are more important than the on-line charges and the off-line printouts. (For example, only accession numbers were pulled in the initial CIS search. It took 3 hours to recover the information for 38 citations from the CIS hardcopy in the library for these numbers.)

Although some of the data bases were low in efficiency and effectiveness it would be a mistake to conduct a more restrictive search. Nearly all the material recovered was discrete; each data base concentrates on a narrow spectrum and overlap is minimal. All the retrievals were important to comprehensiveness. And, the relative usefulness of retrievals was not measured; one particular hit from an otherwise unproductive data base might be worth a dozen others and never be found in a manual search.

The broad approach is a quick way to get into current information on a complex problem. It would be a particularly fruitful, first step for an ad hoc special study group to use. This suggests a new field for librarianship. Such study groups (contract or in-house) need a librarian, experienced in computer-based reference services, as a member of their research team. This librarian must participate fully in the planning of the study, in evaluating the results recovered in the computer-based searches, and in providing his expertise to a multidisciplinary approach. It is questionable whether a librarian in a part-time, ancillary role would become conversant enough with the subject and the progress of the study to contribute as much.

The approach is cost-effective. A librarian using manual search methods would work six days steadily, retrieving and transcribing ten citations an hour, to produce the same results at an equal cost. There are a number of assumptions in that statement, the most important that the librarian using manual methods could actually bring back everything with the same precision as the computer. The statement can be challenged, but even if the librarian, working like a machine, recovered thirty citations an hour, it would take two full days, uninterrupted, a mind-boggling implication. (Appendix 4, Annex C, p. C-4-1, sets forth the rationale.)

The cost-benefit of the searches should be considered from three viewpoints: efficiency (ie, recoveries per unit time), precision (hits per recovery), and effectiveness (hits and near misses per recovery). (See Annex C.)

The approach produces a wealth and variety of information. The sample of retrievals for Question 2 (Annex D) compared with the material on aluminum, above, will show how this method searches all corners of the subject.

The choice of words is important. Learning the terminology is the most critical aspect of the research step (above), and careful attention to potentially key words appearing on-line materially help formulate follow-on searches. In the original search (p. A-10-1) the concept of nationalization was tested in various forms to no avail, while "producer" delivered 28 citations; and, a cluster of event codes revolving around ownership produced 39 in my search (p. A-6-1).

The librarian is going to have to accept the fact that the indexers and abstracters have their own way of looking at words and try to learn them. In Metadex, for example, (p. A-4-1), the words "process" and "refining" retrieved citations pertinent not only to the production of aluminum but also to the use of aluminum in its various forms in other industrial processes. An attempt to exclude the latter, unwanted use of these words would probably exclude useful material as well as consume time on-line. The better practice, given the current state of the art, would be to dump the set off-line for printing and cull unwanted uses manually.

The power of the category codes found in several data bases should be mentioned in connection with this conclusion. They are immensely useful to saving time and increasing precision, but expansion of their use is in the hands of the data base vendors and the computer-based services.

Too narrow a search must be avoided, especially in exploring a large and complex subject. A preplanned course of action should be abandoned once a set is narrowed down to a reasonable size; and, the set should be dumped off-line for printing. An example illustrating this is at Appendix 1, Annex A.

It should be obvious that the librarian's role in the broad approach is quite different from that in a more conventional search. More is demanded of the librarian, who plays a more active role in the direction the search takes. Skimming the surface rapidly of a broad subject in all its complexities he will see things (movements and innovations) that are not apparent to the researcher immersed in the subject. The librarian participating in the broad approach must be readily conversant with the subject to be knowledgeable enough to exploit unexpected retrievals. This can be gained only through long discussions with the researchers, a broad if superficial reading into the subject, and a careful study of the data base literature. This exploration was highly structured. The outer edges of the work could have been performed by a competent operator employing the more conventional techniques associated with a narrow search of a specific topic. The in-depth work would require the full-time participation of a librarian conversant with the subject.

This exploration raises but does not answer the question of how far each should reach out the extra distance required -- the librarian or the client. Both need some familiarity with the other's expertise, but to what extent? Essentially, it's a question of which is easier to do well: grasp a new way of thinking (the organization of indexes and computer-based logic), or assimilate a broad but superficial knowledge of a subject area and its terminology in a discipline, perhaps completely foreign to the librarian's academic background.

The return on the librarian's investment of time in preliminary study as far as the data bases and the computer services go is well worth it for the future. I probably now know more about aluminum than I will ever use again.

The dynamic of the search broadens the questions. Question 2 was originally understood to mean that the economic equation could be transformed if research developed new technology to produce aluminum cheaply from some other mineral besides bauxite. The responses soon made clear (cf. Annex D) that R&D was not limited to this aspect of production cost but all, e.g. energy consumption, so that any advance in the technology of aluminum production could make an increase in the price of bauxite tolerable, or, equally, might lead the major producers to abandoning bauxite entirely, leaving the countries with bauxite with a wasted asset. The question had become broadened to include any research promising to effect overall price and provided an insight that would have been gained only gradually and painfully by manual searching. The searches in GEO-REF disclosing discovery of bauxite in heretofore unsuspected countries affected not only Question 2 but Question 3. (Or, what would happen to OPEC if a few more Alaskan and North Sea oilfields were found?)

The careful study and preparation was a precaution imposed by the expense of time on-line. This constraint paid off not only in expenses but in other ways. Certainly preplanned and written search strategies are recommended. Experience at improvising was dismal. More interesting is where the experience showed inadequate preparation.

Familiarity with Government Reports Announcements and Index would have prepared me for the futility of sampling a large NTIS set which is sequenced alphanumerically by accession number. (NTIS uses the accession numbers of the agencies and contractors; AD identifies the Defense Documentation Center; Defense investigations of aluminum in its many forms is not pertinent to the "big question" of this hypothesis.) Not being aware that PREDICAST had split the file I wanted was an unpleasant surprise, compounded by the fact the split-off did not use all the product codes available in the old. (See Appendix 6, Annex A, p. A-6-3.) The librarian using computer-based reference services

must be aggressive in obtaining and absorbing the changes in the newsletters the services publish. The attempt to fish in World Aluminum Abstracts without a user's manual was a disaster.

This approach will not replace serendipity nor the need for a sensitive nose on the part of the researcher. It should get him quickly into the problem, spending his time absorbing and evaluating information and viewpoints, not plumbing the card files and bibliographic aids. It can change the whole meaning of scholarship in the practical sense.

- Observations made and Lessons learned -

Comparisons between the computer-based reference services are meaningless. One uses the service that has the data bases appropriate to the subject. In this experiment the only data base offered by both the services used was NTIS.

A more useful question is the relationship between the specific data bases and the dimensions of the "big question." It became apparent during the evaluation of the retrievals that the data bases responded to same question at distinctly different levels.

Pursuing this line, a three-dimensional matrix was developed. (See p. B-1-1.) The x-coordinate represents the three questions; the y-coordinate represents the level of treatment, broadly defined as corporate and governmental policy, the production process, and fundamental research. The z-coordinate represents the time dimension; some responses have an immediate bearing, and others are off in the future.

Each data base concerns itself with a particular clientele and a particular range of serials and documents. Metadex, for example, responded to Question 2 with answers ranging from a mathematical model of an alumina reduction cell to environmental policy implications over the entire production stream. NTIS, on the other hand, responded with reports of fundamental research. Comparisons are tricky, and this experiment has insufficient depth to draw conclusions about the data bases. The illustration, however, makes the point -- one gets out of the data bases only what has been put in, and a broad search must use every data base pertinent to the subject.

Congressional Information Service and the Funk and Scott Index (CIS and P18 in the diagram) are very different in both the clientele addressed and the material indexed. These are broad-ranging data bases with catholic interests. Their treatment may not be detailed but they are addressed to policy makers.

More of the recoveries were made in response to Question 2 than either of the others. This is due not only to the broadness of the question but also to the number of data bases related to it. If the client had been interested only in the first or third the search might have been restricted to CIS, ASI, and PREDICAST. But, to have restricted it to just one would have resulted in the same disappointment found in the original, April search. A lot of really pertinent information would have been excluded.

The concept embodied in this matrix ought to be useful in formulating the questions, discussing the subject with the client, and writing the search strategies. Determining the dimensions of the "big question" beforehand and visualizing what data bases will fill what intersections of the various dimensions should clarify thoughts on the subject and on the searches.

There is danger in applying the concept too rigidly. The boxes do not divide retrievals neatly. A policy level response in one data base may be more specific than a fundamental research response in another. The diagram is meant only as an aid in recognizing the multifaceted nature of the relation between the "big question" and the data bases.

The idea of quick-reference tables should be extended to constructing flip charts on a small easel alongside the terminal. The flash cards would contain the connect routines, file call-up commands, and other housekeeping commands related to each computer-based reference service and each data base.

The pressure while on-line is intense. The clock is running off the "dollars." Two people should work together, one mapping the strategy and keeping track of the overall situation while the other operates the terminal and makes the tactical decisions. The former can not maintain the minute-to-minute appraisal of the developing situation, while the latter can not check the written search strategies for effectiveness, refer back to the typed history for set numbers and other information, nor formulate alternate courses of action to keep the momentum or decide to cut losses. Teamwork, with one reading off lengthy, coded commands (e.g., PC = 333410) to the other at the terminal, pays off in accuracy and time.

Deciding when to dump a set off-line for printout is a delicate and expense-related question. Deciding on the amount of detail (or printout format) is another expense trade-off. Such decisions should be made during preparation.

The real cost of search must consider the cost of running down a hardcopy item of the work cited. This computer-based search will be of little practical value unless the client is backed up with a well-organized library with a wealth of holdings, close at hand. When the availability of material is doubtful, plan to dump the entire record, off-line, so that the retrieval can be evaluated properly before incurring the expense of interlibrary loan or travel to a distant library.

The likelihood of an activity being reported in the literature should be considered in determining whether the initial search has likely recovered all in the data base. We would expect the multinationals to publicize governmental actions threatening expropriation (and the Funk and Scott Index contained a lot) but they are likely to be secretive about their own research. The whole question of determining satisfactory ways of estimating whether one has recovered all there is on a question is one that needs study.

Eight of the ten files searched were structured in varying degrees. The convenience, speed, and comfortableness of searching a highly structured data base are no less real for being subjective. The effectiveness (within the limits of this experiment) can be measured.

	Structured					Unstructured			
	H	M	X	Effect.		H	M	X	Effect.
Intnl Abstracts	3	0	0	100%	ASI	18	1	1	90%
Intnl Time Series	32	0	0	100	CIS Q2	25	6	4	88.6
F&S Q3	30	0	0	100	Q1	16	0	4	80
Q2	38	2	1	97.6					
Geo-Ref	32	12	2	95.6					
Metadex	14	8	9	70.9					
					NTIS	5	5	10	50%

There are difficulties with NTIS not experienced with the other data bases. NTIS is structured in a sense, but the controlled vocabularies and other devices are those of the reporting agencies, not NTIS. In effect a miscellany of entries are dumped into a pot with no NTIS-imposed controls other than the COSATI and NTIS category codes, which, in this experiment, at least, were too broad to be effective retrieval aids. Unhappily, accession numbers are no help, either. The accession number of an entry may be that of the governmental agency sponsor, the contractor, or an NTIS-assigned number in default of any other. Success with the Lockheed capability of string search (see Appendix 5, Annex A, p. A-5-2) using "Bureau (lw) Mines" excursion does not promise much help unless the client can supply all the agencies and contracting activities that are doing research in the field under scrutiny.

It might be cost-effective in a particular exploration to dump all of a large set off-line and cull the trash manually. Not in this particular experiment. Much of the NTIS material is at the fundamental research level while the thrust of the hypothesis is at the world economic level.

That the data bases carry anomolies should not be surprising. Some curious ones thrown up in this search were: PREDICAST credited the Soviet Union with producing 5 million tons of bauxite in 1968, although it is known that the USSR mines anorthosite because it has no commercial deposits of bauxite; the estimate must have developed a bauxite equivalent. The F&S Index reported MIT was speculating on asteroid mining for aluminum, although the actual article in Industrial Research did not mention aluminum at all. Even indexers and abstracters make mistakes.

Even though the on-line data bases do not run back through the entire history of the hardcopy sources, the retrievals give good clues as to which journals are particularly fruitful in a given area. The researchers, if they are so inclined, can pursue the thread as far as they wish in the hardcopy indexes.

A number of experiments suggested themselves for other explorers to investigate. One mentioned is how to determine what part of the relevant material has been retrieved. Another would be to explore a subject such as gerontology in all its ramifications -- social, medical, economic -- in an entirely different mix of data bases and services to see what can be added to the concept of a broad search.

Case studies in searches are lacking for library schools and other activities teaching computer-based reference services.

Should the broad approach be limited to data bases with highly structured codes and other aids and the loosely structured data bases be searched manually? The evidence found in this exploration indicates otherwise but it is not conclusive.

If there is to be a real exploitation of this approach the services must improve the documentation they supply. The high-cost of the services will keep even the most active users at a low level of familiarity with each data base. The manuals must supply the difference or few will dare to use an unfamiliar data base in a broad approach.

Improvements in data base structure and the services are also needed. The on-line data bases are essentially the same as their hardcopy versions except for the very important addition of the inverted indexes created by tearing apart titles, abstracts, etc for free search terms. A couple of hundred controlled words and/or category codes seem hardly adequate to reflect all the facets of

a base with 100,000 entries, growing at a rate of 10,000 a month. The controlled vocabularies of today are mere lists of "okay" words until scope notes, related terms, and other syndetic measures are introduced. They need to be called up on-line, too, and more use made of "invalid option" responses to help the user not totally familiar with the data base. (See my experience with PREDICAST noted above.)

Nuances are needed which deal with document content and orientation. More sophisticated sampling techniques are required to deal with large sets before dumping them off-line to print. Taking the first ten in descending order of accession number won't cut it.

In some data bases certain index terms are in a condition of permanent overflow. The services cannot abandon the user with such unmanageable globs. After all the reason such terms are in overflow is because they are so common and necessary to the subject discipline.

Librarians, vendors, and the services need to work together to develop this approach as well as the narrow search. The potential is truly enormous. Its successful development can revolutionize what is thought of as scholarly research. It can be of immeasurable help to business making critical decisions in the dark.

When Lewis and Clark returned from the Pacific their report raised more questions than it could answer. Certainly this short exploration of mine can, at best, intrigue others and prompt them to experiment with the broad approach. The potential of computer-based reference service has barely been touched. It promises to be an unimaginably large bibliography that can be entered on a multiplicity of axes quickly to retrieve any kind of information the client wants.

ANNEX A

SEARCH HISTORIES

The on-line record of the searches conducted on 27 October, 3 and 11 November 1977 are contained in this annex. Glitches that were caused by typographical errors in commands, transients in the lines, and false starts have been eliminated to simplify following the logic used and the results obtained. Wherever a set was printed on-line an asterisk (*) is used. Where only a sample was taken for test, the number taken is shown in brackets. Bracket statements are placed in the text to explain category codes so the reader need not look them up; bracket notes in Appendix 8 following set size are for the convenience of the reader to show the amount of overlap obtained in fishing for a substitute for the unknown subject categories desired.

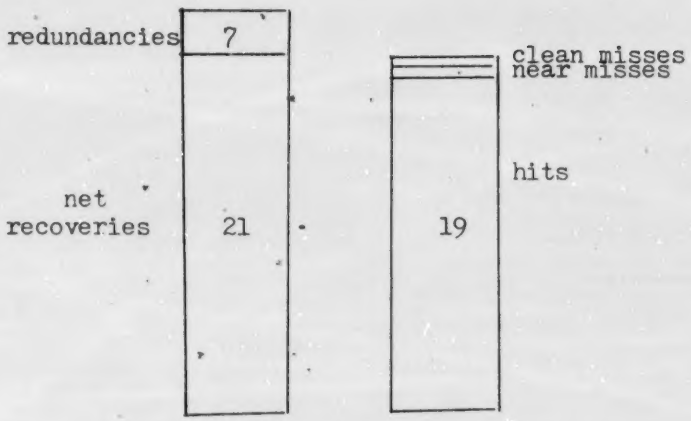
The bar charts at the foot of each search show the recoveries made (net and redundancies) in the left-hand column and the hits, near misses, and clean misses in the right. (These charts are not to the same scale from one appendix or page to the other.)

Statistics regarding the searches will be found in Annex C, in tabular and graphical form. Descriptions of the data bases used will be found in Annex B.

Appendix 1 ASI (American Statistical Index)
Appendix 2 CIS (Congressional Information Service)
Appendix 3 GEO-REF (American Geological Institute)
Appendix 4 METADEX (Metals Abstracts and Index)
Appendix 5 NTIS (National Technological Information Service)
Appendix 6 PTS (PREDICAST Terminal System)
Dialog .file 18: F&S Index
Dialog .file 86: International Statistics Abstracts
Dialog .file 87: International Annual Time Series
Appendix 7 SSIE (Smithsonian Science Information Exchange)
Appendix 8 WAA (World Aluminum Abstracts)
Appendix 9 Summary of time on line of the searches
Appendix 10 Original search, conducted by client, 18 April 1977

ASI -- Question 1

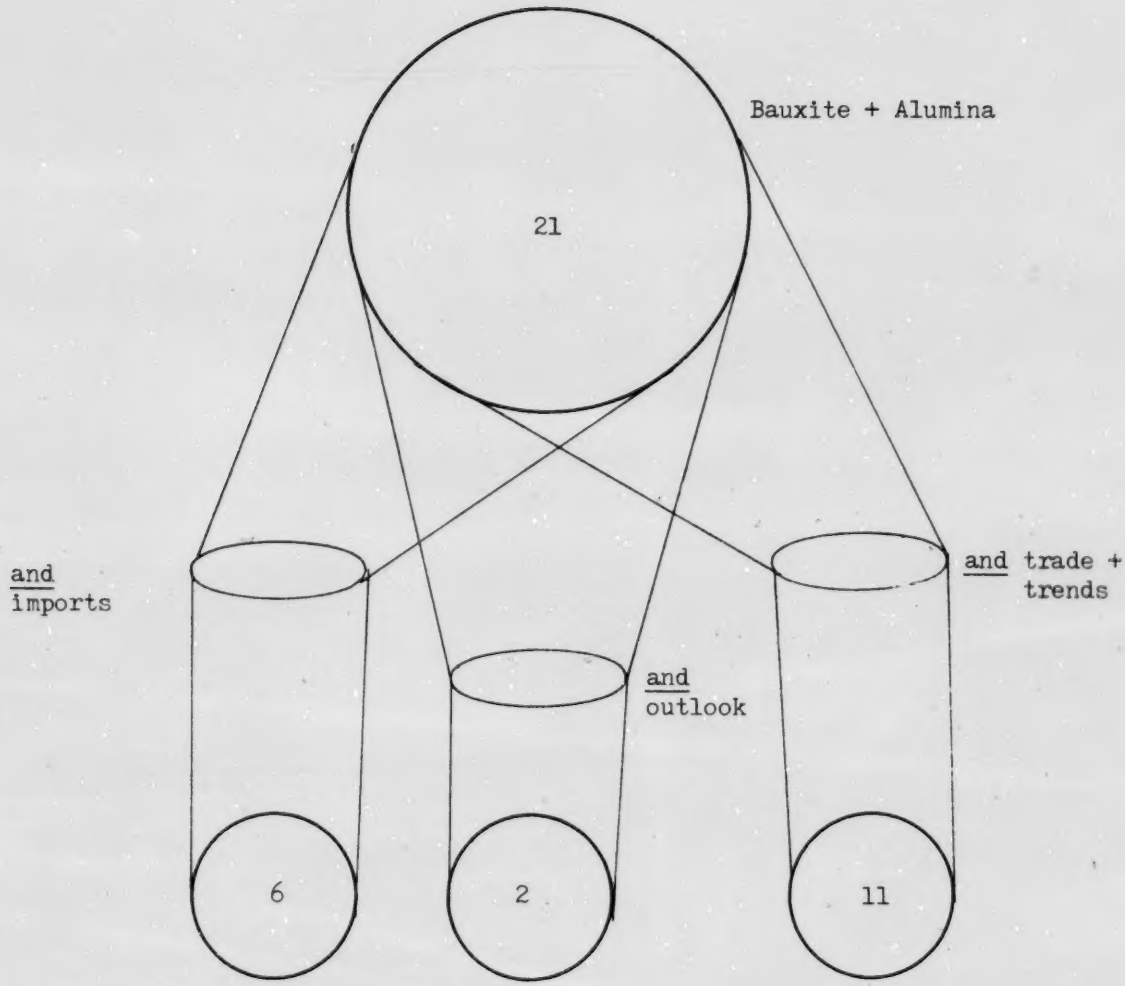
All alumina or all bauxite	SS1	21
1 and imports	SS2	6*
1 and outlook	SS3	2*
1 and trade or 1 and trends	SS4	11*
2 or 3 or 4	SS5	12
1 and not 5	SS6	9*



SEARCH RESULTS

ASI -- QUESTION 1

Initial Search



RESTRICTIVE EFFECTS OF VARIOUS QUALIFYING AND'S

CIS -- Question 1

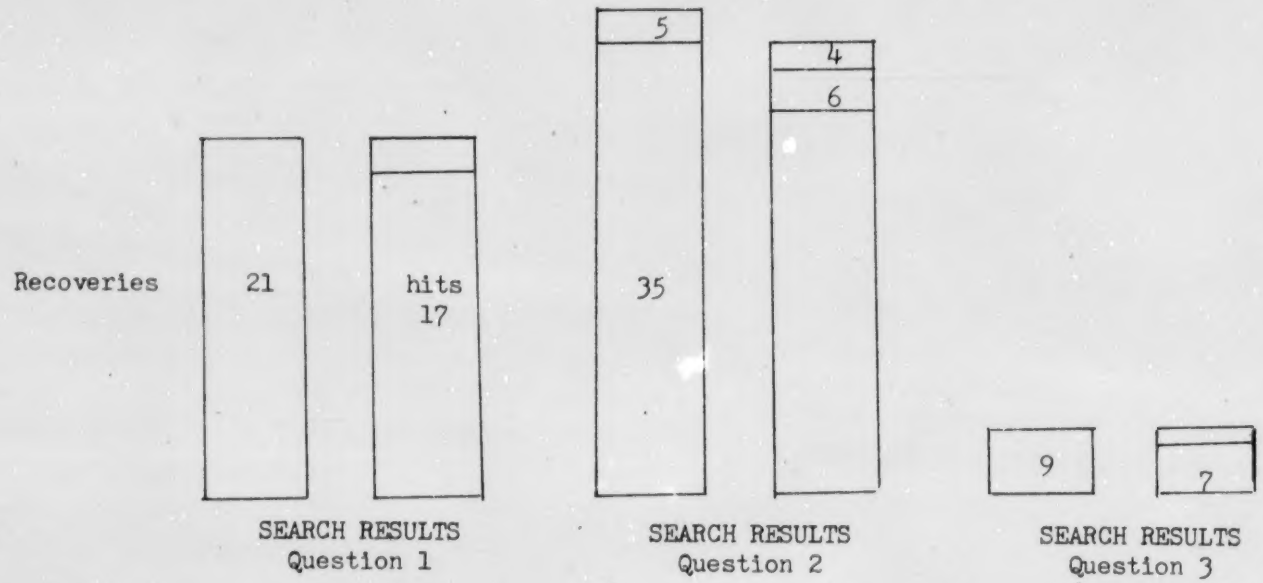
Alumina (FT) or bauxite (FT)	SS1	20
All import	SS2	643
All import:	SS3	3094
1 and 3	SS4	7 *
1 and not 4	SS5	14 *

CIS -- Question 2

All aluminum or all bauxite or all alumina	SS6	81
All industry or all mining or all production	SS7	5618
6 and 7	SS8	56
ø (DT) or 1 (DT) or 2 (DT) or 3 (DT)	SS9	25775
9 and 8	SS10	55
All technolog: or all process: or all techni:	SS11	4472
10 and 11	SS12	30
12 and all new	SS13	9 *

CIS -- QUESTION 3

Acqui: (FT) or Expropria: (FT) or Cartel: (FT)	SS14	1323
14 and 9	SS15	1260
15 and 6	SS16	9 *

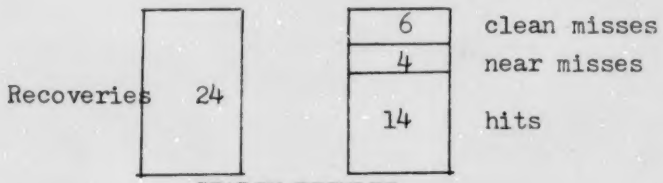


CIS -- QUESTION 2

A set of 30 narrowed to 9 was printed out during the initial search; this search reruns the initial commands and recovers the others in that set to compare and determine whether the initial recovery was comprehensive.

In addition it uses some terms turned up in the first search to explore the data base in an additional direction.

All alumina or all aluminum or all bauxite	SS1	81
All industry or all mining or all production	SS2	5618
1 and 2	SS3	56
All technolog: or all process: or all techni:	SS4	4472
3 and 4	SS5	30
5 and all new	SS6	9
5 and not 6	SS7	24*
Resource: (FT) or need: (FT) or scarcit: (FT)	SS8	8012
All raw (FT) and all material (FT)	SS9	39
8 or 9	SS10	8019
10 and 1	SS11	42
11 and not 7	SS12	31*



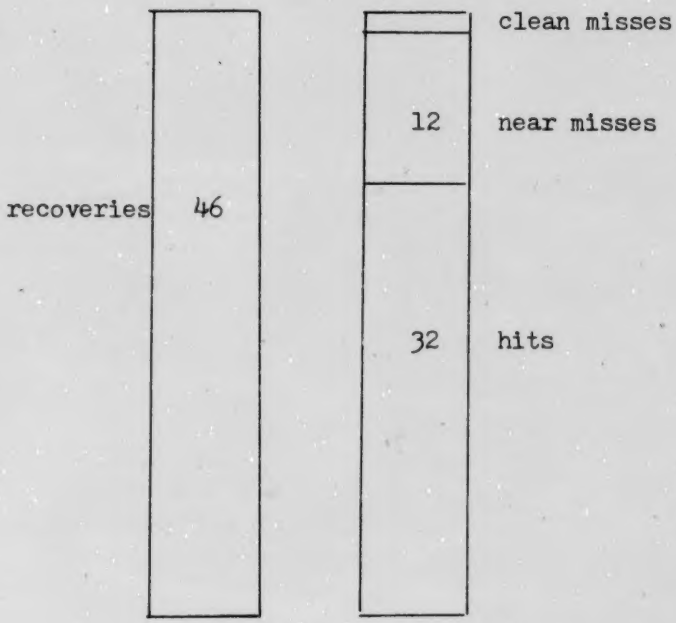
SEARCH RESULTS

Excursion

Removing new

GEO-REF -- Question 1

All alumina: and EL(LA) or all bauxit: and EL(LA)	SS1	448
1 and from 76 thru 77	SS2	45
Discover: or explor:	SS3	14682
2 and 3	SS4	3*
1 and 3	SS5	46
5 and not 4	SS6	43*

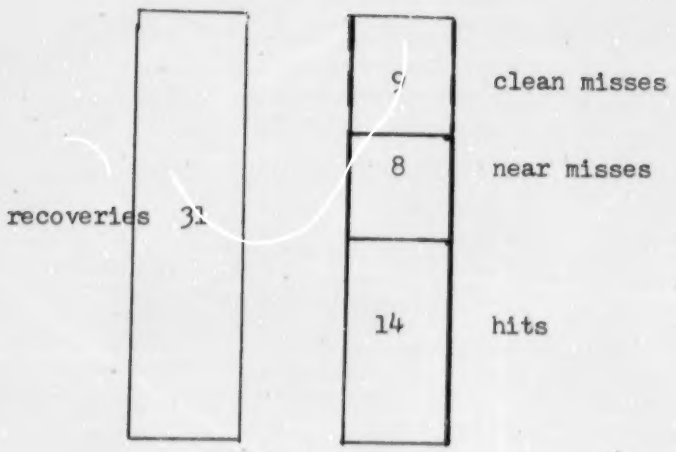


SEARCH RESULTS

METADEX -- Question 2

S SC=41 [ores & minerals]	29	1083
S SC=42 [extracting & smelting]	30	5295
S SC=43 [refining & purification]	32	2513
C 29-30/or	33	8891
S alumin?	34	36287
C 33 and 34	35	963
S alumina	36	646
C 33 and 36	37	108*

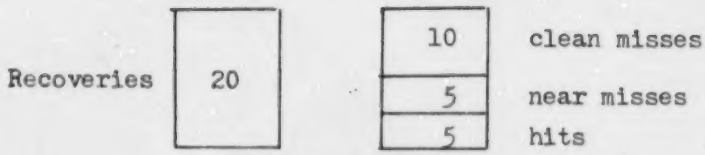
[output cut off at 31, when entry reached back to 1974]



SEARCH RESULTS

NTIS/DIALOG -- Question 2

S aluminum	1	11112
S alumina	2	1088
S bauxite	3	31
S CF=11F [metallurgy & metallography]	8	16258
C 1 and 8 [aluminum & metallurgy & metallography]	9	3408
C 2 and 8	10	110
C 3 and 8	11	5*
S CF=13H [industrial processes]	12	7248
C 2 and 12 [alumina & industrial processes]	13	105 [sample 5]
S CF=96C [international commerce & marketing]	14	1325
C 2 and 14 [alumina & internl commerce....]	15	0
C 1 and 14 [aluminum & internl commerce...]	16	7*
C 3 and 14 [bauxite & internl commerce....]	17	1
C 16 and 17	18	1
S CF=Ø8I [mining engineeering]	19	1097
C 3 and 19 [bauxite & mining engineering]	20	3*
S new (w) developments	21	67
C 3 and 21 [bauxite & new developments]	22	0
C 1 and 21 [aluminum & new developments]	23	3
C 2 and 21 [alumina & new developments]	24	0



SEARCH RESULTS

NTIS/DIALOG -- Question 2

Recovery from the initial search was disappointing; it seemed that NTIS should offer more. The initial search concentrated on bauxite, the mineral which is input to the aluminum-producing process. This follow-on, an excursion, concentrated on alumina, the second stage in the process.

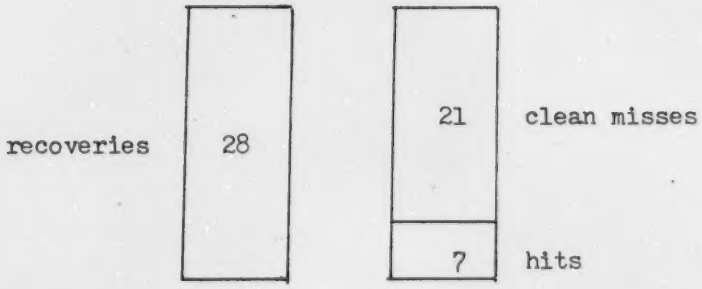
S alumina	1	1088
S CF=11F [metallurgy & metallography]	2	16258
C 1 and 2	3	110 [sample 10]
S Bureau (lw) Mines	4	2523
C 3 and 4	5	7*

See table, p. C-2-1 for results

NTIS/ORBIT -- Question 2

NTIS was the only data base used which is found in both the Lockheed (DIALOG) and SDC (ORBIT) services. Therefore, this search, an excursion, went into the NTIS/ORBIT data base, duplicating the searches made in NTIS/DIALOG, for comparison.

All alumina or all aluminum or all bauxite	SS1	6810
11F (CC) [metallurgy & metallography]	SS2	11036
1 and 2	SS3	2226
3 and Ø8I [mining engineering]	SS4	18 [sample 5]
All alumina or all bauxite	SS5	707
5 and 2 [metallurgy & metallography]	SS6	64
6 and Ø8I (CC)	SS7	4*
6 and 13H (CC) [industrial processes]	SS8	4
8 and not 7	SS9	4*
6 and not 8 or 6 and not 9	SS10	60
6 and not 8	SS11	60
6 and not 9	SS12	60
11 and 12	SS13	60
6 and not 7	SS14	60
11 and 14	SS15	56 [sample 15]



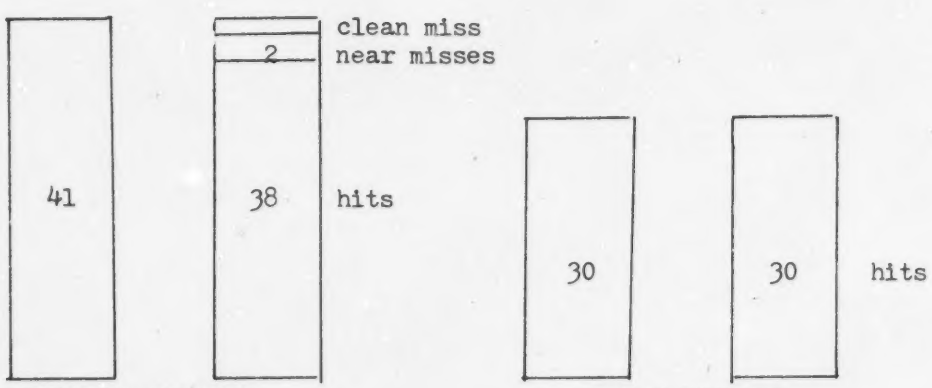
SEARCH RESULTS

F&S INDEX (P18) -- Question 2

S PC=3334000 [aluminum]	2	4324
S PC=1051000 [bauxite & alumina]	4	208
S EC=310 [science & research]	5	162
S EC=314 [applied research]	6	355
S EC=318 [engineering procedures]	7	109
S EC=321 [process development]	8	2457
S EC=323 [process introduction]	9	3898
C 5-9/or	10	6094
C 2 and 10	11	39*
C 4 and 10	12	2*

F&S INDEX -- Question 3

S EC=126 [expropriation]	13	2052
S EC=146 [foreign operations]	14	6175
S EC=148 [parent-to-joint venture]	15	9575
S EC=150 [acquisition & merger]	16	1529
S EC=952 [tariffs & quotas]	17	3530
S EC=955 [foreign investment regulation]	18	791
S EC=958 [export controls]	19	1314
S EC=962 [nationalism]	20	189
S EC=960 [international politics]	21	366
C 13-21/or	22	24881
C 2 or 4 [aluminum or alumina & bauxite]	23	4500
C 22 and 23	24	288 [sample 5]
C 22 and 4 [alumina & bauxite]	25	39*



INTERNATIONAL STATISTICAL ABSTRACTS (P86) -- Question 1

S PC=1051000 [bauxite & alumina]	38	16
S EC= 412 [mineral reserves]	39	31
S EC=462 [use of minerals]	40	0
S MC=30 [tons]	41	52295
C 38 and 39 and 41 [mineral reserves*bauxite & alumina*tons]	42	0
C 38 and 39	43	0
C 38 and 41 [bauxite & alumina*tons]	44	3*
C 39 and 41 [mineral reserves*tons]	45	8

See table, p. C-1-1 for results

INTERNATIONAL ANNUAL TIME SERIES (P87) -- Question 1

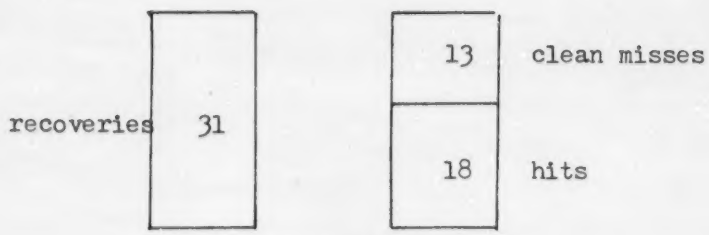
The results of searching International Statistical Abstracts (P86) were disappointing. Seeking outside advice it was learned that PTS File 2, International Statistics, (as described in the PTS Users Manual available), had been redesignated and subdivided. The follow-on search was therefore directed to the International Annual Time Series (DIALOG File 87).

S PC=1051000 [bauxite & alumina]	43	0
S PN=bauxite and alum	44	0
S PC=1051100 [bauxite]	45	48
S PC=1051200 [alumina]	46	8
C 45 and 46	47	0
C 45 or 46	48	56
S MC=30 [tons]	49	32399
C 48 and 49	51	32
S SF=BASEBOOK	52	85982
C 51 and 52	53	32*

See table, p. C-1-1 for results.

SSIE -- Question 2

All alumin: and all materials	SS1	671
All produc: or all refin:	SS2	58180
1 and 2	SS3	105
Alumina (ST) and 2	SS4	31*



SEARCH RESULTS

WAA -- Question 2

The literature available in the computer access laboratory, GSL, DU, contained nothing on World Aluminum Abstracts. Information in Computer-Readable Bibliographic Data Bases, however, suggested that WAA might be productive, since it was produced by the American Metals Society, producer of METADEX, and it was focused on aluminum ore processing, exclusive of mining. An assumption was made, proven false by results and subsequent checking with a library subscribing to WAA that the subject category codes (SC) would be the same as in METADEX. An attempt was made, on-line, to call up a description of WAA SC's, but the expand command response was merely the SC numbers in proximity. The attempt, in the latter part of this search, to call up responses by combining up words associated with the desired SC's was successful.

S SC=41 [assumed to be ores & raw materials]	6	2782
S SC=42 [assumed to be extracting & smelting]	7	805
S SC=43 [assumed to be refining & purification]	8	937
C 6-8/or	9	4504
S alumin?	11	31670
C 9 and 11	12	2131
S alumina	15	1987
C 9 and 15	16	34 [sample 10]
C 6 or 7 [assumed to be ores & raw materials, extracting & smelting]	17	3567
C 17 and 15	18	7*
S ores	19	335
S raw	20	403
S materials	21	7774
C 19-21/or	22	8202 [only 310 in common]
C 22 and alumin?	23	4111
S extract?	24	1322
S smelt?	25	367
C 24 or 25	26	1656 [only 33 in common]
C 24 and alumin?	27	1327
C 23 and 27	28	298 [sample 5]

See table, p. C-2-1 for results.

TIME ON LINE OF THE SEARCHES

ASI - first connection	.24 hours
ASI - excursion	.06
CIS - first connection	.49
CIS - second connection	.46
GEO-REF - first connection	.23
METADEX - first connection	.178
NTIS/L - first connection	.326
NTIS/L - second connection	.134
NTIS/SDC - excursion	.27
PTS .file 18 - first connection	.535
PTS .file 86 - first connection	.071
PTS .file 87 - second connection	.31
SSIE - first connection	.40
WAA - first connection	.26
	3.964 hours

F&S INDEX (P18)

S PC=E41210 [mining production]	1	75
S PC=105100 [bauxite & alumina]	2	189
S PC=106200 [manganese ores]	3	210
S PC=106911 [chromium ores]	4	220
S PC=109940 [tin ores]	5	401
S PC=333410 [primary aluminum]	6	507
S PC=333440 [refined aluminum]	7	97
S PC=333976 [tin]	8	616
S PC=331244 [tin mill production]	9	59
S EC=962 [nationalism]	10	89
S EC=150 [merger data]	11	1838
S EC=110 [name]	12	24
S nationalization	13	399
S nationalizatn	14	11
S nationalizes/nationalizing	15	0
C nationalization and nationalizat	16	3
S nationalize	17	140
S nationalized	18	216
S nationalizes	19	88
S nationalizing	20	19
C nationalization and nationalizat	21	407
C 17-20/or	22	446
C 2-9/or	23	2207
C 1 and 23	24	0
C nationalization or nationalizat	25	407
C 22 or 25	26	843
C 26 and 23	27	2*
S production	28	41708
S producers	29	2164
S exporters	30	259
S exportation	31	16
C 28-31/or	32	2436
C 32 and 23	33	28 [printed off-line]
C 11 or 12 [merger data + name]	34	1862
C 34 and 33	35	0

Recoveries	28
Hits	27
Clean misses	1
Useful	14

ANNEX B

THE DIALOG AND ORBIT DATA BASES USED - A VIEWPOINT

There were some 45 data bases available at DU at the time of this experiment, October-November 1977. Nine were selected for consideration; one ABI/Inform was dropped because its orientation did not promise much application to this problem. There was no information available on one, World Aluminum Abstracts, and the difficulties with working under such circumstances is discussed in the main body of the report.

Although some of the 8 data bases used were low in efficiency and effectiveness for this experiment it can be no measure of the usefulness of their retrievals to the client. One citation can be of as much value as all the rest, together. Each data base has a relatively narrow focus, concentrating on a particular direction; the overlap was minimal.

A concept evolved from an increasing awareness of the points made above as the individual retrievals were transcribed and examined. The concept is shown graphically in Appendix 1 to this annex. The client's problem had been broken down into 3 questions which form the x-axis of the visualization. The retrievals responded, roughly, to three levels of interest: corporate and national policy, the total production process, and fundamental research. These are represented on the y-axis. Some responses address the immediate problems of the aluminum industry; others are more appropriate to the near and distant future. These form the z-axis.

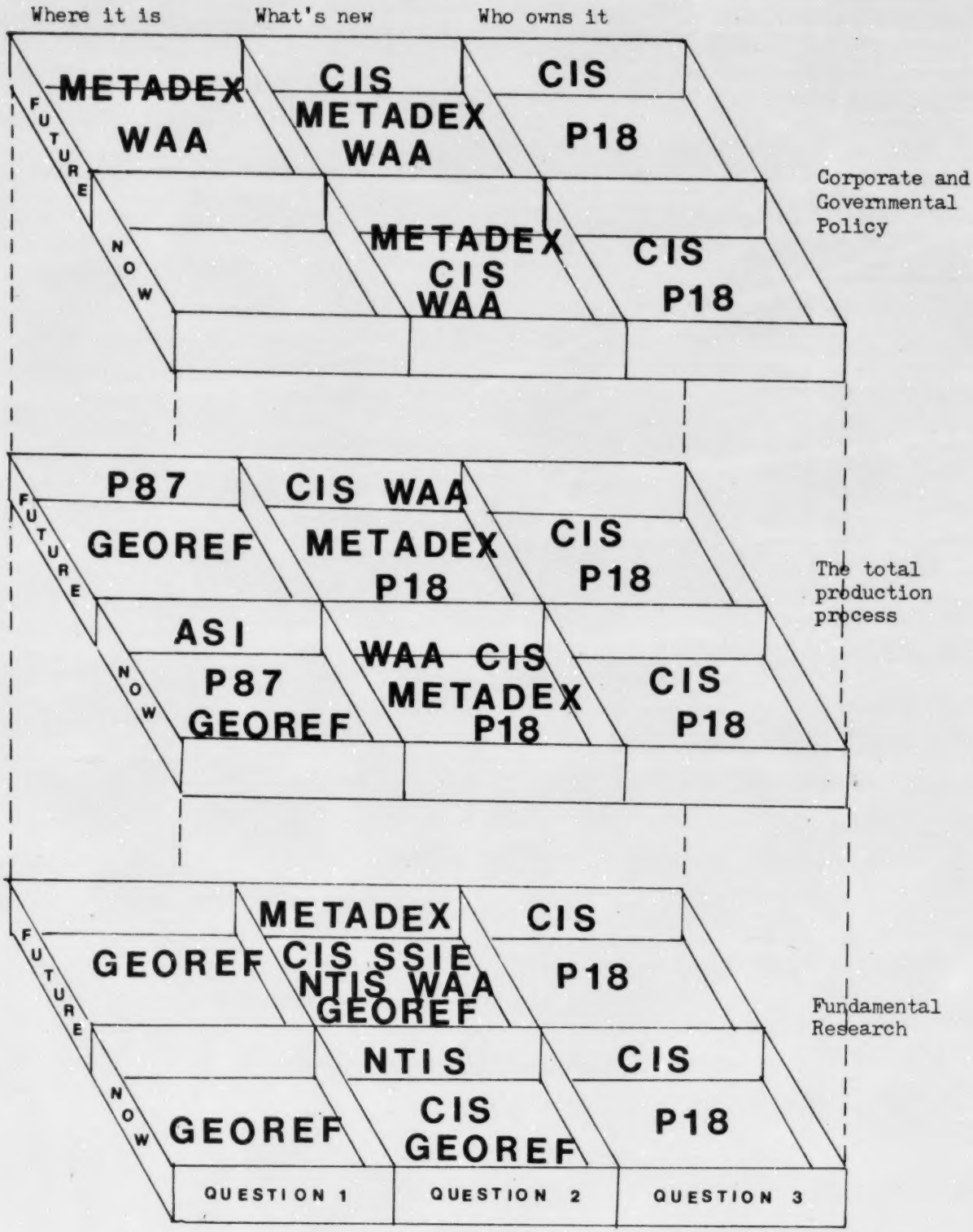
The pertinence of the responses of the data bases in this experiment are roughly shown by putting their abbreviation in the appropriate boxes. (P18, P86, and P87 are used for appropriate PREDICAST files to avoid clutter.) The boundaries of the boxes in the matrix are too hard and fast properly to depict the relationship, but it will have to do. I lack skill in perspective drawing and airbrush technique to represent it properly with spheres intersecting, all circumscribed by a great sphere representing the big question -- the client's problem.

Another dimension of the data bases from the viewpoint of this experiment is their level of detail or specificity which I have attempted to depict with the scalar diagram in Appendix 2.

In order that this paper may stand alone (and the reader need not consult other references) a brief description of the purpose and scope of each of the data bases used is included in Appendix 2.

I reached no conclusions about the data bases themselves and tried to exclude such considerations in the work. I could not however help but think that NTIS is a difficult data base to work with because of the way it is assembled, using whatever structure the particular governmental agencies use within their own literature, and imposing no integrating structure of its own.

An arbitrary but reasonable standard of 75% precision was established to lead to further consideration of why a particular data base did not meet it in this experiment. A discussion is found in Appendix 3 to the annex.



PERTINENCE OF THE DATA BASES TO THE DIMENSIONS OF THE CLIENT'S PROBLEM

DESCRIPTION OF THE DATA BASES

ASI (American Statistics Index): a master guide to the statistical publications of the United States Government, both documents containing tabular statistics and narrative material about statistics or consisting of statistical information. Its purpose is to provide a single source in which to locate all data currently collected by the Federal Government on a particular subject.

CIS (Congressional Information Service Index): provides access to a wide body of information emanating from the work of the 300 odd committees and subcommittees of the U.S. Congress.

GEO-REF (American Geological Institute): covers some 29 different areas in the field of the geosciences. Those of particular interest to this experiment were economic geology (general and mining) and economic geology-metals, but applied geophysics retrieval proved valuable. GEO-REF entries come principally from journal articles, but about 25% are from non-serial documents.

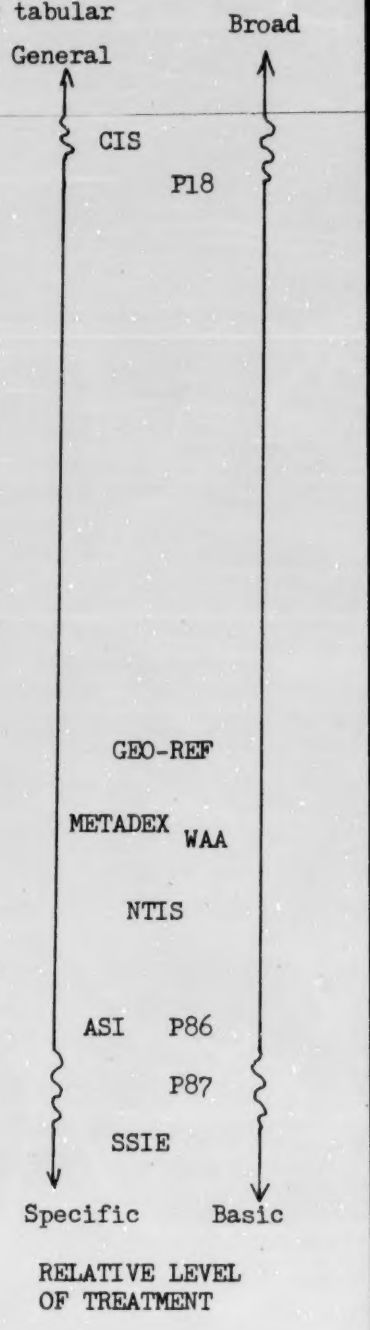
METADEX (Metal Abstracts and Index): is designed for researchers and scientests working in areas related to metallurgy; 80% of the material is drawn from journal articles, in English and in other languages.

NTIS (National Technological Information Service): contains citations and abstracts of Government-sponsored research and development reports and other analyses prepared by Federal agencies or their grantees or contractors. References are those announced in NTIS's Weekly Government Abstracts and Government Reports Announcements and Index.

P18 (PREDICAST's F&S Index - DIALOG .file 18): briefs articles relevant to business research from newspapers, trade journals, government plans, bank letters, prospectuses and reports. All products, services, events, industries, companies and countries are covered.

P86 (PREDICAST's International Statistics Abstracts - Dialog .file 86): abstracts of published forecasts on general economics, all industries, detailed products and end-use data.

P87 (PREDICAST's International Annual Time Series - DIALOG .file 87): annual data for production, consumption, price, foreign trade and usage statistics.



SSIE (Smithsonian Science Information Exchange): prepublication research information from federal, state, local government, associations, foundations, individual investigators, universities, and industry.

WAA (World Aluminum Abstracts): for researchers, scientests, and managers in aluminum ore processing (less mining) through end uses.

CONSIDERATION OF THE PRECISION OBTAINED

The terms used, precision, efficiency, and effectiveness, as used in this report are elsewhere defined. Precision is the number of hits over the net recoveries yielded by a particular search. This appendix considers why certain data bases yielded precisions below 75% in this experiment, only.

Question 1 Recoveries

ASI	P=.857
CIS	.8095
GEO-REF	.696
P86	1.000
P87	1.000

Question 2 Recoveries

CIS	.714
SSIE	.58
NTIS/L	.25
P18	.927
METADEX	.452

Question 3 Recoveries

CIS	.778
P18	1.000

GEO-REF: was searched with free terms using the algorithm: alumina or bauxite and discover or explore, restricted to English language documents. The retrievals included citations on the means of exploring or discovering, the relationship between geological phenomena and exploration and discovery, and policy matters related to exploration and discovery as well as information on the discovery or exploration of specific deposits.

SSIE: was searched with the subject term (alumina) and the modifiers, production or refining. The retrievals included research projects in ceramics, dentistry, pharmacology, and jewelry (the sapphire is a pure crystalline alumina) as well as in the process of producing aluminum from alumina and alumina from bauxite.

NTIS/L: the samples from the large sets created were output in descending order of accession numbers; these were in alphanumeric order and the majority sampled were from the Defense Documentation Center (AD) whose interests are not related to the client's problem.

METADEX: was searched with subject codes for ores & minerals, extracting & smelting, and refining & purification. Because of METADEX's level of treatment more than half dealt with the basic physics and chemistries of the processes, which were not pertinent to the client's problem -- no payoff in altering the total equation within the time frame he is addressing.

The standard, P less than 75%, was chosen a priori to examination of the results. Except for rigidly constructed vocabularies and codes, a broad search should yield some misses if not drawn too tight. Examination of results bear out the reasonableness of the standard.

ANNEX C

DESCRIPTIVE STATISTICS OF THE RESULTS

This experiment was an exploration of a new and broad approach to the use of computer-based reference services. There was no intent and no work done to pursue any line of inquiry exhaustively -- to prove anything. Excursions were made down interesting trails; searches were cut off when they were unproductive and then abandoned. There was neither the time nor intent to find out whether a particular data base was appropriate; for example, WAA, which promised a lot after the structure of its subject categories was learned, was never explored.

Descriptive statistics have been compiled and are presented in this annex to illuminate further the information presented on the searches.

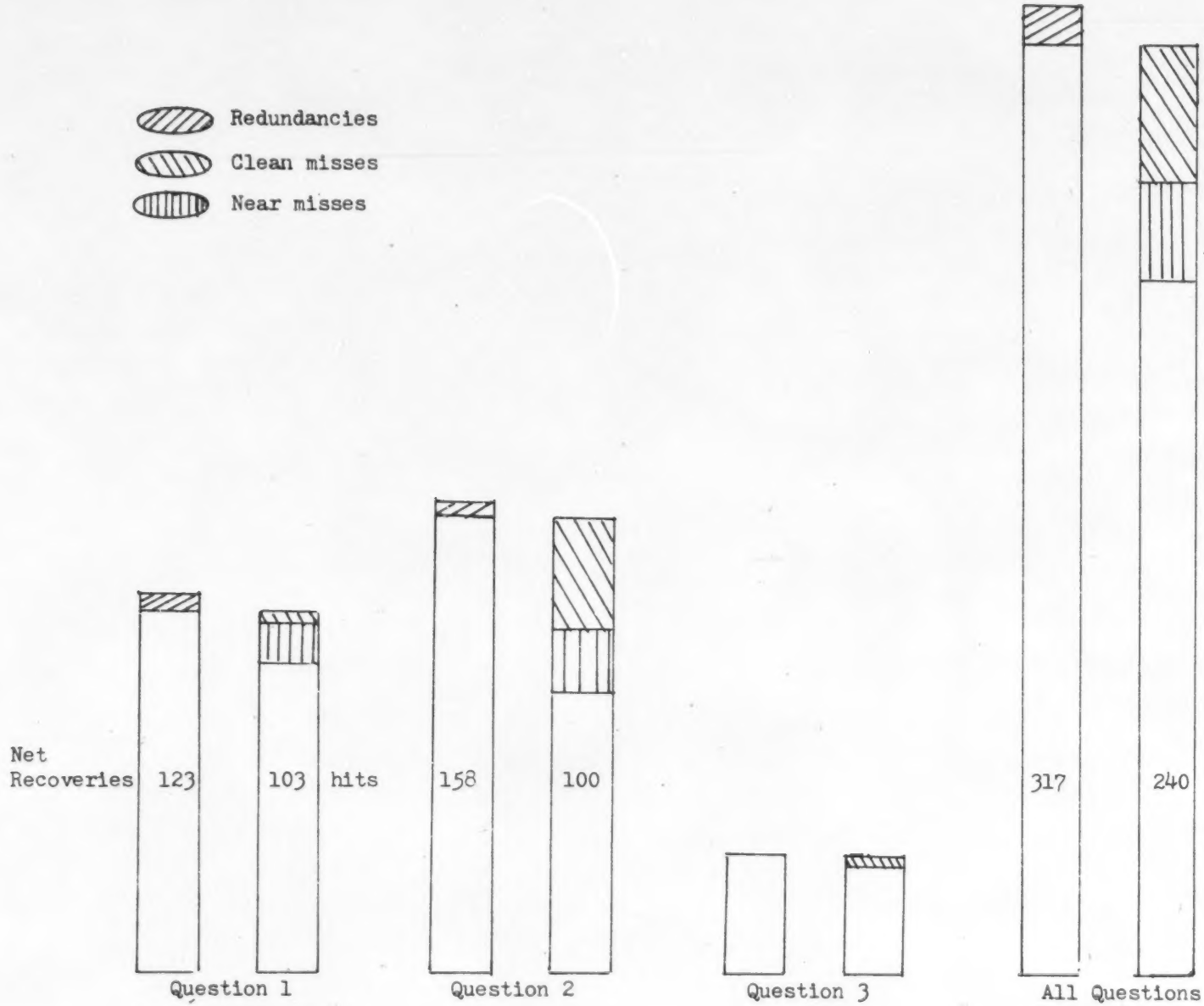
Appendix 1 is a table showing the statistical results of the primary searches. Appendix 2 contains the same information on the excursions. Appendix 3 tabulates the efficiency, effectiveness, and precision of the various searches. Appendix 4 is a cost comparison of the manual with the computer-based search.

STATISTICAL RESULTS-PRIMARY SEARCHES

Data Base	Recoveries	Redundancies	Net Recoveries	Hits	Near Misses	Clean Misses
ASI	28	7	21	19	1	1
CIS	21	0	21	17	0	4
GEO-REF	46	0	46	32	12	2
P86	3	0	3	3	0	0
P87	32	0	32	32	0	0
Sum, Ques 1	130	7	123	103	13	7
CIS[1]	40	5	35	25	6	4
SSIE	31	0	31	18	0	13
NTIS/L	20	0	20	5	5	10
P18	41	0	41	38	2	1
METADEX	31	0	31	14	8	9
Sum, Ques 2	163	5	158	100	21	37
CIS	9	0	9	7	0	2
P18	30	0	30	30	0	0
Sum, Ques 3	39	0	39	37	0	2
Sum, All Ques	332	12	320	240	34	46
Efficiency	2.17/min	3.9%	96%	72.3%	10.3%	13.9%
Orig Search	28	0	28	27	0	1
Efficiency	1.78/min	0%	100%	96.4%	0	3.6%

Note 1: Broadening the search in the follow-on using terms generated by the initial search increased the recovery from 9 to 40, hits from 7 to 30 (23 additional), with only 5 redundancies, and an increase of 1 to 6 near misses, and 1 to 4 clean misses. The conclusion -- to return to data bases being searched with free terms to broaden it -- is apparent.

GRAPHICAL DISPLAY OF THE PRIMARY SEARCHES



STATISTICAL RESULTS-EXCURSIONS

Data Base	Recoveries	Redundancies	Net Recoveries	Hits	Near Misses	Clean Misses
Question 2:						
CIS[2]	24	0	24	14	4	6
NTIS/L[3] Alumina	10	0	10	0	0	10
NTIS/L BuMines	7	0	7	5	0	2
WAA SC41-43	10	0	10	0	0	10
WAA SC41-42	7	0	7	0	0	7
WAA free terms	5	0	5	4	0	1
NTIS/SDC	28	0	28	7	0	21

Notes:

2. Excursion to determine how restricting the free term, new, was. Omitting it enlarged the recoveries from 9 to 33 (gain of 24), hits from 7 to 21 (gain of 14), near misses from 1 to 5 (gain of 4), and clean misses from 1 to 7, (gain of 6).

3. Separate excursions were conducted, using alumina instead of bauxite, and using Bureau (lw) Mines. The first set of 110 were sampled. All 10 of the sample were from Defense Documentation Center (accession numbers "AD+six digits) and purpose was defeated. The Bureau of Mines approach was much more productive, but searching by agency would require the client or librarian to have a grasp of who was publishing what that probably would be impractical.

TABLES OF EFFICIENCY, PRECISION, AND EFFECTIVENESS

Search	Efficiency	Precision	Effectiveness
Question 1:			
ASI	20 in .24 hrs 1.38/min	18/20 90%	19/20 95%
CIS	*	17/21 80.95%	17/21 80.95%
GEO-REF	46 in .23 hrs 3.33/min	32/46 69.5%	44/46 95.7%
P86	3 in .071 hrs .7/min	3/3 100%	3/3 100%
P87	32 in .31 hrs 1.72/min	32/32 100%	32/32 100%
Question 2:			
CIS	*	25/35 71.4%	31/35 88.6%
SSIE	31 in .4 hrs 1.29/min	18/31 58%	18/31 58%
NTIS/L	20 in .326 hrs 1.02/min	5/20 25%	10/20 50%
P18	**	38/41 92.7%	40/41 97.6%
METADEX	31 in .178 hrs 2.9/min	14/31 45.2%	22/31 70.96%
Question 3:			
CIS	*	7/9 77.7%	7/9 77.7%
P18	**	30/30 100%	30/30 100%

* Efficiency of CIS overall - 65 in .49 hrs = 2.21/min
** Efficiency of P18 overall - 71 in .535 hrs = 2.21/min

Efficiency	
GEO-REF	3.33/min
METADEX	2.90/min
CIS	2.21/min
P18	
P87	1.72/min
ASI	1.38/min
SSIE	1.29/min
NTIS/L	1.02/min
P86	.70/min

Effectiveness	
P18 (Q3)	100%
P86	100%
P87	100%
P18 (Q2)	97.6%
GEO-REF	95.7%
ASI	95%
CIS (Q2)	88.6%
CIS (Q1)	80.95%
CIS (Q3)	77.7%
METADEX	70.96%
SSIE	58%
NTIS/L	50%

Cost Comparison - Man and Machine

1. Librarian at $ 12,000 per annum.

52 weeks at 40 hours/week =	2080
2 weeks vacation	80
1 week sick leave	40
2, 20 minute breaks/day	163.3
16% fatigue factor	287.5
useful time/year	1509.2

pay + 6% FICA tax = $ 12,720
cost per hour = 8.43

2. Assume an on-line cost of $ 100/hour
   Librarian's cost 8.43
   Cost per hour CBR search 108.43

3. Time on-line, experiment's primary searches = 2.54
   Assumed cost, time-charges, experiment = $ 275.41
   Cost per retrieval: 275.41/332 = $ 0.83

4. Equal cost time required (t) for manual searching, retrieval, and transcription by the librarian:

83¢ = 332/8.43 t; t = 47.45 hrs, or 5.93 days at the rate of 10 (8.43/.83) retrievals per hour.

5. Excluded from consideration are the cost of fringe benefits (hospitalization, etc) and the time necessary for librarian to interview the client and become sufficiently oriented to conduct the search effectively.

ANNEX D

An example of the wealth and diversity of information recovered in a broad search as shown in a sample of titles retrieved for Question 2.

Electrical utility rates...energy information...meeting America's resource needs...mineral scarcity...energy waste and energy efficiency...materials shortages: paper and aluminum...outlook for prices and supplies of industrial raw materials...multinational corporations and U.S. foreign policy...recycling waste materials...stockpile commodities...disposal of aluminum from the national stock pile...aluminum and sodium minerals from oil shale...ALCOA buys land containing a possible bauxite substitute...aluminum producers to begin to use nonbauxitic ores on a commercial basis...inert gas refining of aluminum cuts gas consumption ...Barclays assesses nonferrous metals situation...new process for aluminum metal production by chloride electrolysis...40% energy saving in new production process...alternate aluminum producing methods advance slowly...alunite studied as a substitute for bauxite...new plant converts coal waste to alumina... "revolutionary" new process for removing aluminum from clay...process to recover aluminum from fly ash of coal combustion...alumina from kaolin...alumina from Utah alunite...control of red mud concentration...production of alumina by pyrolysis...aluminum oxide from bauxites and wastes...removal of iron from aluminum ores...gallium recovery from the Bayer process [bauxite to alumina]...optimum water management in kaolin mining...removal of contaminants during disposal of sludge...production of alumina from domestic sources...conserving carbon in electrodes in refining...solvent extraction technology in alumina production... the activities of AlF_3 and NaF in melts of saturated alumina... .

ANNEX E

BAUXITE PRODUCERS

Country	Production (millions of tons) in 1974	Growth Rate 1960-1974	% World Production in 1974
Australia	18.54	64.4%	23.5%
Jamaica	15.22	6.3	19.3
Guinea	7.605	9.6	9.6
Surinam	6.863	5.9	8.7
Guyana	3.613	2.1	4.5
France	2.938	3.3	3.7
Greece	2.813	7.6	3.6
Hungary	2.751	6	3.5
USA	2.408	1.8	3
Yugoslavia	2.37	5.4	3
Indonesia	1.29	9.5	1.6
India	1.113	9	1.4
Dominican Republic	1.119	4	1.4
Malayasia	.947	3.9	1.2
Brazil	.9	16	1.1
Romania	.816	24.6	1
Haiti	.793	9.6	1
Sierra Leone	.650	26.3	
Ghana	.357	4.2	
Italy	.032	-14.9	
Spain	.009	3.6	
Mozambique	.005	0	
Rhodesia	.002	6.5	
West Germany	.001	-9.1	
USSR	4.3 Estimated	-0.4	5.4
China	.6 Estimated	3.6	0.7
WORLD	78.81	9.2	

Source: P87 output

ANNEX F

ALUMINUM COMPANIES

The majors

Alcan Aluminum Limited
Pechiney Ugine Kuhlman Corp.
Alcoa (Aluminum Company of America)
Reynolds Metals Company
Kaiser Aluminum and Chemical Company
Schweizerisches Aluminum AG (Alasuisse)

Important minors

Revere Copper and Brass
Anaconda Aluminum Company
Martin-Marietta Aluminum
Howmet Corporation
Olin Corporation